心理学报 2018, Vol. 50, No.3, 317−326
Acta Psychologica Sinica DOI: 10.3724/SP.J.1041.2018.00317

普遍信任博弈决策的动态过程 ——来自脑电时频分析的证据*

付 超[1,2] 张 振[1] 何金洲[3] 黄四林[4] 仇剑崟[5] 王益文[1,2]

([1] 福州大学心理与认知科学研究所, 福州 350116) ([2] 福州大学经济与管理学院, 福州 350116)
([3] 中原大学心理系, 桃园 32023) ([4] 北京师范大学发展心理研究院, 北京 100875)
([5] 上海交通大学医学院附属精神卫生中心, 上海 200030)

摘 要 人类社会的特殊性在于普遍信任贯穿于所有的人际互动, 人们为什么以及何时选择信任的问题对社会发展和经济交易具有重要意义。然而, 现有研究结果尚无法清楚地阐述信任决策的认知与神经机制。为探究信任互动情境下大脑活动变化的时间动态特征及神经震荡过程, 采用事件相关电位技术记录了21名健康被试作为信任者完成单次性信任博弈(Trust Game, TG)时的脑电数据, 并对EEG数据进行时域和频域分析。行为结果发现个体选择信任的比例显著高于几率水平。时域分析结果表明决策阶段中不信任决策比信任决策诱发了更负的N2成分(250~330 ms), 更小的P3波幅(330~430 ms)。频域分析结果则发现不信任决策比信任决策引发了更大的β频段(18~28 Hz, 250~350 ms)能量活动; 相比于获益反馈后信任决策, 损失反馈后信任决策诱发了更大的β频段(15~20 Hz, 150~300 ms)能量活动, 而损失反馈后不信任决策和获益反馈后不信任决策诱发的β频段能量无显著差异。时域N2成分和频域β频段能量的动态分离表明信任选择是人们的默认优势选项, 而不信任选择则涉及更多的认知控制加工。本研究为理解人类信任决策过程的动态加工时程及其神经震荡特征提供了初步的电生理学证据。

关键词 普遍信任; 决策; N2; P3; 时频分析
分类号 B849:C91

1 引言

普遍信任(Generalized Trust)是指个体在普遍意义上对他人(包括那些素未相识的陌生人)的一种相对抽象的态度与行为(Zak & Knack, 2001), 它孕育着人类社会的社会合作与稳定性。普遍信任是人类社会互动过程中一个必要成分, 能够降低社会交易成本, 易化合作行为, 促进社会经济的繁荣发展, 长久以来被视为合作行为的一个关键前提(Krueger et al., 2007)。实验研究经常采用反映现实情境核心特征的博弈任务, 探究与分析人类的信任行为(Berg, Dickhaut, & McCabe, 1995; Declerck, Boone, & Emonds, 2013; Krueger, Grafman, & McCabe, 2008; Tzieropoulos, 2013)。信任博弈(Trust Game, TG)是研究人类信任行为的一个经典博弈范式, 往往由两名匿名玩家共同完成, 双方分别拥有一定数额的金钱S, 要求一名玩家(信任者)把部分金钱Y ($0 \le Y \le S$)交给另一名玩家(受托者), 然后受托者获得3Y的金钱, 并决定返还给信任者X ($0 \le X \le 3Y$)的金钱, 最终信任者收益为$S-Y+X$, 受托者收益为$S+3Y-X$。大量研究表明信任博弈中存在着较普遍的合作行为, 即信任者会将部分金钱交给对方, 而大多数受托者也会返还一定比例的金钱(Camerer, 2003; Johnson & Mislin, 2011)。

收稿日期: 2016-10-23
* 国家自然科学基金项目(31771238, 31371045)、福建省闽江学者支持计划、福建省社会科学规划项目(FJ2017C049)资助。
张振与付超为共同第一作者。
通信作者: 王益文, E-mail: wangeven@126.com; 仇剑崟, E-mail: jianyin_qiu@163.com

随着脑成像技术的发展，研究者开始逐步探究信任行为背后的认知神经机制。一系列的功能性磁共振成像技术(functional-Magnetic Resonance Imaging, fMRI)研究表明，信任者做出信任选择时会显著激活内侧前额叶皮层(medial prefrontal cortex, mPFC)，尾状核(caudate nucleus)，杏仁核(amygdala)，脑岛(insula)和背侧前额叶皮层(dorsolateral prefrontal cortex, dlPFC)等脑区(Aimone, Houser, & Weber, 2014; Baumgartner, Heinrichs, Vonlanthen, Fischbacher, & Fehr, 2008; King-Casas et al., 2005; Knoch, Schneider, Schunk, Hohmann, & Fehr, 2009; Krueger et al., 2008; McCabe, Houser, Ryan, Smith, & Trouard, 2001)。研究者认为 mPFC 主要参与心理推理、情绪认知与调控等过程，尾状核参与同伴值得信赖性的习得与编码，杏仁核和脑岛则主要参与不确定性和背叛恐惧加工，而 dlPFC 则参与认知控制过程(Declerck et al., 2013; Riedl & Javor, 2012)。相较而言，目前有关信任时间进程的研究则比较匮乏，且大多聚焦于信任决策的结果评价过程(Chen et al., 2012; Long, Jiang, & Zhou, 2012; Ma, Meng, & Shen, 2015)。例如, Ma 等(2015)发现事先承诺可以促进信任行为，且事先承诺条件下非奖赏反馈减奖赏反馈引起更大的 FRN，表明事先承诺条件下被试具有更强烈的互惠预期。有鉴于此，当前研究拟采用 ERP 技术和单次性信任博弈任务，探究信任博弈决策过程中大脑活动的时间动态特征。

对于博弈决策过程而言，研究者所关注的核心问题是“个体为什么会信任或不信任他人(尤其是陌生人)，两种选项何者为优势反应?”。虽然大量研究结果表明人类存在较普遍的信任行为，但是研究者对上述议题的理解还存在强烈的争论。背叛厌恶理论(betrayal aversion theory)强调信任困境涉及到一种策略不确定性，总是伴随着遭受对方背叛或剥削的可能性，并且人们存在普遍的背叛厌恶动机(即回避遭受背叛所导致的负性情绪的欲望)，因此人们倾向于选择不信任进而表现出较少的信任行为(Bohnet & Zeckhauser, 2004, Bohnet, Greig, Herrmann, & Zeckhauser, 2008)。有研究发现个体在信任博弈中的投资额度或比例要显著低于赌博任务，而且这种差异可能与前脑岛的激活有关(Aimone et al., 2014)。道德规范理论(moral norm theory)则认为信任是一种由内化的道德性强制规范驱动的行为，个体借此向他人品格表示尊重，维持其关于他人值得信赖与诚信善意的社会矫饰，进而表现出较高的信任行为(Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014; Dunning, Fetchenhauer, & Schlösser, 2012)。Dunning 等(2014)发现人们会过分的信任陌生人，即使这种信任水平超过了其风险忍受性，并且与社会责任履行有关的情绪(如内疚和焦虑)能够部分程度上解释这种行为。据研究所知，目前只有极少数 ERP 研究为博弈决策的时间进程提供了初步的实证数据(Boudreau, McCubbins, & Doulson, 2009; Wang et al., 2015)。Boudreau 等(2009)采用抛硬币任务操纵了个体与匿名报告者利益冲突、利益一致和撒谎惩罚三种条件，发现利益一致条件下报告者的声明信息诱发了更大的 P2、P3 和晚期正成分。一项采用重复性信任博弈的脑电研究则发现不信任决策比信任决策诱发了更正的 P2 成分，差异波溯源分析定位于额中回和额下回，可能反映了决策过程中信息整合引起的冲突检测加工(Wang et al., 2015)。尽管上述研究已初步探究了博弈决策的时间进程，但也存在一些不足之处。一方面，两项研究所采用的实验任务均有一定的局限性，其中抛硬币任务所操纵的人际信任与传统信任定义不一致，无法准确反映信任互动的特性；而重复信任博弈侧重于互动同伴值得信赖性或声誉信息的习得与编码，无法探究个体在普遍意义上对陌生他人的态度与行为。另一方面，认知控制相关研究普遍认为 N2 是反映认知控制过程的核心成分，它是一种具有额中央分布的负性成分，对冲突检测和反应抑制特别敏感，表现为高冲突情境要比低冲突情境能诱发更负的 N2 成分(Nieuwenhuis, Yeung, van den Wildenberg, & Ridderinkhof, 2003; van Veen & Carter, 2002; Schmajuk, Liotti, Busse, & Woldorff, 2006)。现有信任博弈研究没有发现不同决策导致的 N2 成分差异，对信任或不信任选择何者为优势反应的问题尚未做出清晰回答。因此，当前研究旨在以 N2 成分为指标，进一步深入探究博弈决策中不同选择的优势性问题。

另外，现有研究大多只关注时域上锁时锁相的事件相关电位(event-related potential, ERP)，而非相位锁定的神经震荡(event-related oscillations, ERO)则反映了与刺激呈现锁时的 EEG 节律能量变化；这些 EEG 活动的能量变化反映了刺激所诱发的神经元节律性的调整过程(兴奋或抑制)，并且不同频段的神经震荡往往与人类的各种认知活动有关(Tiesinga, Fellous, & Sejnowski, 2008; Wang, 2010)。认知控制相关的神经震荡研究发现抑制反应能够

导致右侧额下回等脑区上 β 频段(13~30 Hz)能量增加, 表明 β 震荡能够调节行为抑制潜在的神经网络活动(Alegre et al., 2004; Krämer, Knight, & Münte, 2011; Swann et al., 2009)。相比于停止信号任务(Stop-signal Task)中的反应试次和失败的抑制试次, 成功的抑制试次导致了更强的 β 频段能量(Marco-Pallarés, Camara, Münte, & Rodríguez-Fornells, 2008)。一项采用 go/no-go 范式的研究显示, go 试次诱发了额中央区域 β 频段能量的减弱, 而 no-go 试次则导致了额叶区域 β 频段能量的增加, 进而表明 β 频段的能量变化反映了额叶抑制加工(Alegre et al., 2004)。

现有研究虽然在时域上初步揭示了信任决策的时间动态进程, 但尚未清晰地回答博弈决策中不同选择的优势性问题, 而且在频域分析上仍缺乏对博弈决策的神经震荡机制的探讨。为了更深入地考察信任决策加工的时间动态过程和神经震荡机制, 记录了 21 名健康被试(充当信任者)完成单次性信任博弈时的脑电数据, 并对 EEG 数据进行时域和频域分析。鉴于研究者关于博弈决策中不同行为选择的优势性问题尚存争议, 本研究提出假设: 如果决策博弈过程中信任选择为优势选项(与道德规范理论相一致), 则个体会更多的选择信任, 不信任选择比信任选择诱发更大的 N2 成分和 β 频段能量; 反之, 如果不信任选择为优势选项(与背叛厌恶理论相一致), 则个体会更多的选择不信任, 信任选择比不信任选择诱发更大的 N2 成分和 β 频段能量。另外, 有研究结果表明先前结果效价能够调控个体随后的行为决策(Gu, Ge, Jiang, & Luo, 2010; Masaki, Takeuchi, Gehring, Takasawa, & Yamazaki, 2006), 因此我们依据信任选择后反馈效价(Gain vs. Loss)和随后试次的决策类型(Trust vs. Distrust)构建了 4 种条件并进行频域分析, 以探究先前结果对随后行为决策的影响。

2　方法

2.1　被试

被试为 21 名在校大学生或研究生, 11 男 10 女, 年龄范围为 18~26 岁, 平均年龄为 21.2 岁。所有被试均为右利手, 视力或矫正视力正常, 无精神病史, 且在实验前都签署了知情同意书。依据被试在博弈任务中的最终收益, 实验结束后给予一定报酬。

2.2　任务和程序

当前实验任务为依据 Berg 等人(1995)的投资/信任游戏改编的信任博弈游戏, 其中被试扮演信任者角色与其他玩家完成 150 回合的游戏, 其间记录其脑电活动。在每轮游戏开始之前, 信任者和受托者均会得到 10 点的出场费, 信任者(脑电被试)需要决定是否将所有的 10 点交给受托者(互动同伴)。假如信任者选择不交给受托者, 则当前回合游戏终止, 双方获得各自的 10 点出场费; 假如信任者选择交给受托者, 则全部 10 点将翻 3 倍交给受托者, 并由受托者决定双方如何分配其所有点数(原有 10 点出场费加上翻倍后的 30 点, 共计 40 点), 受托者可以选择独吞所有点数或者选择与信任者平分所有点数, 无论其做出何种决策, 当前回合游戏终止且双方获得相应的点数。

当前信任博弈采用单次匿名互动方式进行, 也就是说每轮游戏中的互动对象都不相同。具体言之, 脑电被试被告知脑电实验前研究人员已从社会上筛选了 400 名具有社会代表性的成人被试, 并要求他们表明“在单次匿名互动情况下, 假定其被信任且获得了 30 点, 那么他们会选择平分或还是独吞所有点数?”, 然后将 400 名被试的选择存贮于计算机当中; 在脑电被试完成的每轮游戏中, 电脑会从中随机选择一个被试的决策以完成此轮游戏。但实际上, 所有受托者的决策都是由实验前编写的电脑程序控制的, 以保证整个博弈中的结果强化率(信任选择后正性反馈的比例)接近 50%。

在脑电被试进入实验室后, 实验者向其介绍信任博弈的详细规则。然后, 被试舒适地坐在电屏蔽室的椅子上, 距离电脑屏幕 1 m 左右, 并进行佩戴电极帽等准备工作。正式实验中被试需要完成 150 回合的信任游戏。每回合开始都先呈现一张信任游戏的简易决策树(1500 ms), 提示被试当前任务的所有可能选项及结果。在一个持续 500 ms 的十字后, 电脑屏幕上呈现一个决策选项图(2000 ms); 被试要在决策选项图呈现时间内按键做出信任或不信任的决定, 选择信任按 1 键, 选择不信任按 3 键, 按键方式在被试间平衡, 超出 2000 ms 视为无效数据。随后, 在随机黑屏(800~1200 ms)后, 屏幕上分别呈现当前回合的结果(1200 ms), 以及目前的回合数与积累的总收益(2000 ms)。

2.3　脑电记录与分析

采用国际 10~20 系统扩展的 64 导电极帽, 以 NeuroScan 系统记录 EEG 信号。脑电记录时所有电极参考置于右乳突的一只参考电极, 离线分析时再次以置于左乳突的有效电极进行再参考, 即从各导联信号中减去 1/2 该参考电极所记录的信号, 转化

为以双侧乳突的平均值为参考。同时记录双眼外侧的水平眼电(HEOG)和左眼上下眶的垂直眼电(VEOG)。滤波带通为 0.05~100 Hz, AC 采样, 采样频率为 1000 Hz/导, 所有电极与头皮之间阻抗都小于 10 kΩ。对数据进行离线分析, 矫正眼电伪迹, 自动排除其他波幅大于± 80 μV 的伪迹信号。为确保信任条件与不信任条件下叠加试次和信噪比的一致性, 对于每名被试而言, 我们以试次数量较小的条件为基准, 从另一种条件中随机抽取相同数量的试次进行叠加, 最终两种条件下的平均有效试次为 46 ± 19 次(Wiswede, Münte, & Rüsseler, 2009)。ERP 分析时程为决策选项图呈现前 200 ms (作为基线)到呈现后 600 ms, 叠加被试选择信任和选择不信任两种条件下的 ERPs 波形图, 最后平均所有被试两种条件下的 ERPs 得到总平均图。依据已有文献(Boudreau et al., 2009; van Veen, & Carter, 2002; Wang et al., 2015)和总平均图的观察与分析(见图 1A), 我们考察决策阶段内的 P2、N2 和 P3 三种脑电成分, 其中 P2 选择测量决策选项呈现后 150~250 ms 内最大峰值, N2 选择测量决策选项呈现后 250~330 ms 内最负峰值, P3 选择测量决策选项呈现后 330~430 ms 内最大峰值。三种脑电成分均进行 2(决策类型: Trust vs. Distrust)×5(电极: Fz、FCz、Cz、CPz、Pz)的重复测量方差分析, 统计结果非球形性时采用 Greenhouse-Geisser 法校正 p 值, 多重比较采用 Bonferroni 法校正。

2.4 时频分析

重新截取决策选项前 1000 ms 和之后 2000 ms 的数据存贮为 EEG 数据, 最终信任条件和不信任条件纳入时频分析的平均有效试次均为 46 ± 19 次。将 EEG 数据采样率降为 500 Hz, 然后采用 EEGLAB 工具箱中一种复杂的正弦小波转换程序对 EEG 数据进行神经震荡分析(Delorme & Makeig, 2004)。这种分析技术运用正弦小波与时域数据卷积, 随着频率升高小波周期也升高, 进而获得对 3~35 Hz 范围内每个时间点上的瞬时震荡能量估计。这种程序先对 EEG 数据进行单试次分析, 再完成多试次的平均, 最终获得各条件下的震荡能量; 震荡能量数值以决策选项前 400~200 ms 为基线进行矫正, 并转换为分贝量尺[10×log (μV^2)]。依据 EEGLAB 工具箱导出事件相关谱扰动(event-related spectral perturbation, ERSP)数据的统计差异图, 视觉观测找到存在显著差异的频段和时间窗口, 然后导出相应范围内的数据均值进行统计分析。当前实验主要聚焦于信任决策阶段中的 β 频段(18~28 Hz, 250~350 ms)。为了简明起见, ERSP 数据选择了 Cz 点作频谱图, 有差异的频段/时程内的全脑数据作地形分布图(见图 2A)。

另外, 我们依据信任选择后反馈效价(Gain vs.

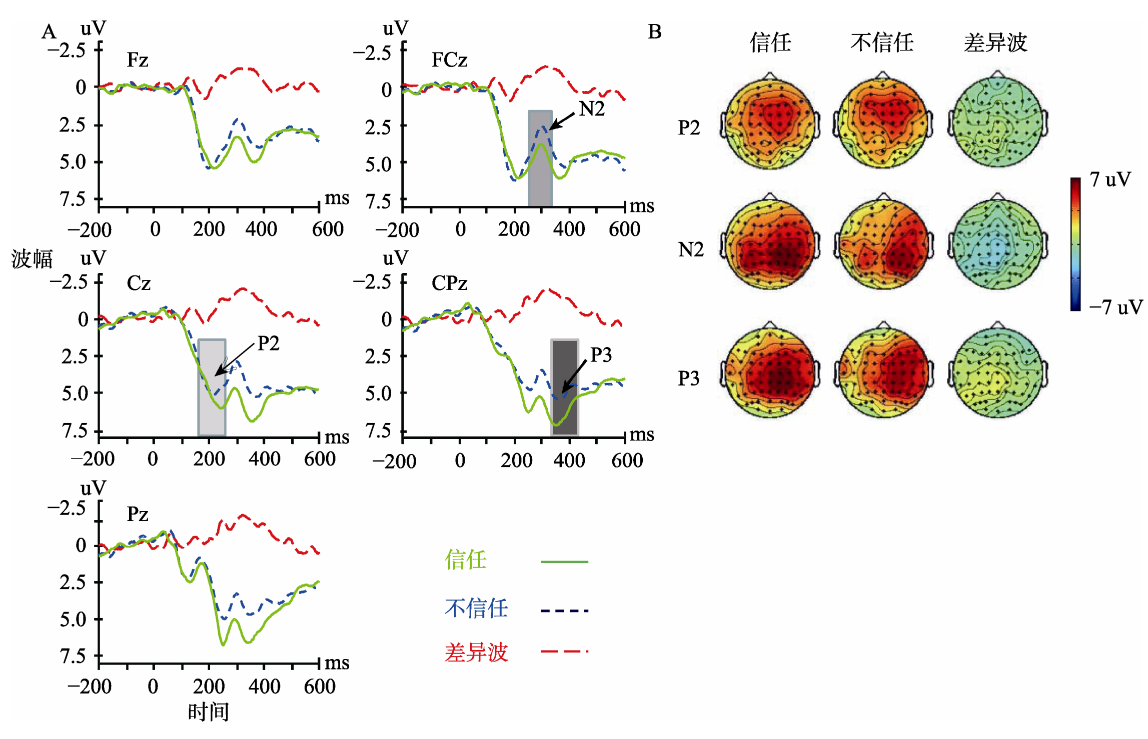


图 1 不同决策类型诱发的 ERP 波形图(A)和 P2 (150~250ms)、N2 (250~330ms)、P3 (330~430ms)的脑地形图(B)

注：彩图见电子版, 下同

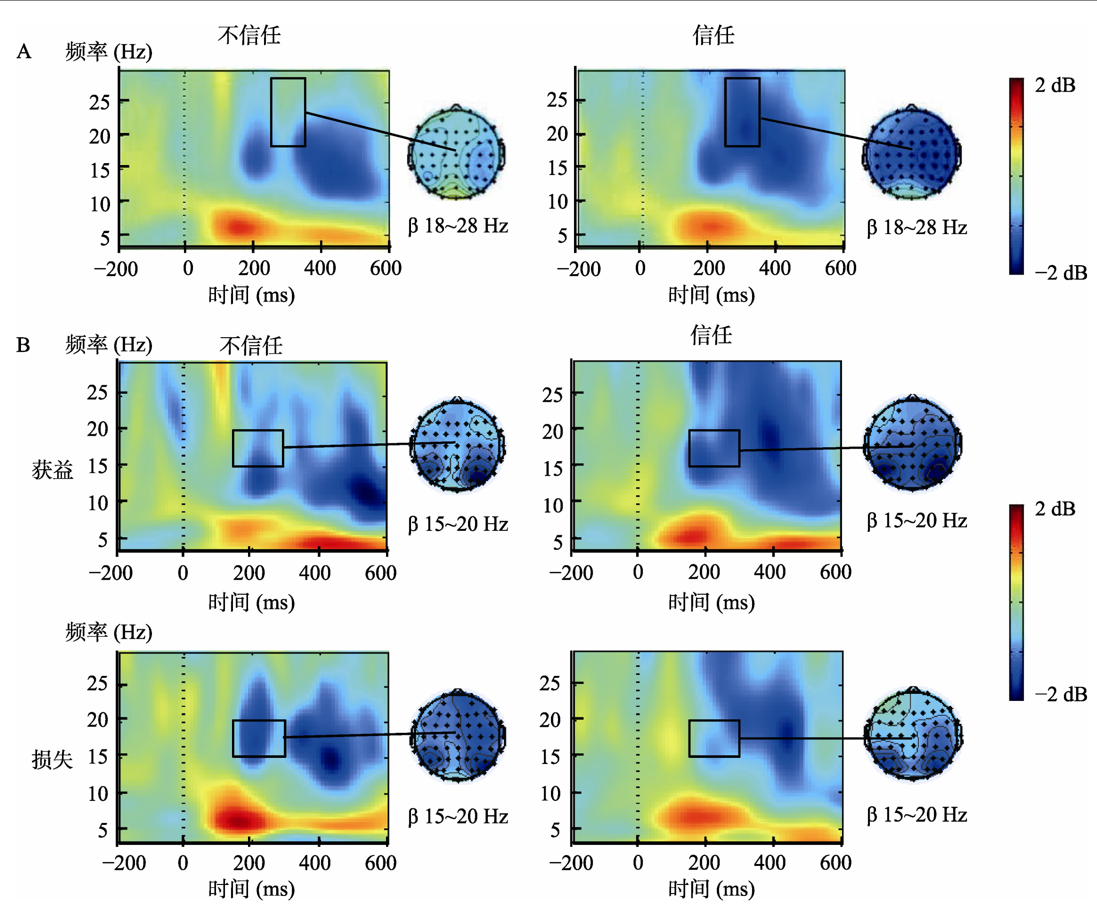


图 2　Cz 点上两种决策类型诱发的 ERSP (A)和不同反馈后两种决策类型诱发的 ERSP (B); 黑色矩形框对应所分析频段的 ERSP, 虚竖线对应决策选项的起始点, 方形图右侧为所分析频段 ERSP 的头皮地形图

Loss)和随后试次决策类型(Trust vs. Distrust)形成了 4 种条件, 分别截取其决策选项前 1000 ms 和之后 2000 ms 的数据存贮为 EEG 数据, 进而试图检验信任选择后反馈效价对随后行为决策的影响。最终获益反馈后不信任决策条件、获益反馈后信任决策条件、损失反馈后不信任决策条件和损失反馈后信任决策条件纳入时频分析的平均有效试次数目分别为 18 ± 8、33 ± 17、13 ± 5 和 34 ± 14。最后, 将 EEG 数据采样率降为 500 Hz, 借助 EEGLAB 工具箱进行相同的神经震荡分析(Delorme & Makeig, 2004)。依据 ERSP 数据的统计差异图, 我们主要关注不同反馈效价后决策阶段中 β 频段(15~20 Hz, 150~300 ms)。为了简明起见, ERSP 数据选择了 Cz 点作频谱图, 有差异的频段/时程内的全脑数据作地形分布图(见图 2B)。对决策类型诱发的 β 频段均进行 2(决策类型: Trust vs. Distrust)×5(电极: Fz、FCz、Cz、CPz、Pz)的重复测量方差分析; 对不同反馈效价后随后行为决策诱发的 β 频段均进行 2(反馈效价: Gain vs. Loss)×2(决策类型: Trust vs. Distrust)×5(电极: Fz、FCz、Cz、CPz、Pz)的重复测量方差分析, 统计结果非球形性时采用 Greenhouse-Geisser 法校正 p 值, 多重比较采用 Bonferroni 法校正。

3　结果

3.1　行为结果

被试选择信任和不信任的比例分别为 67.6 ± 3.1%和 32.4 ± 3.1%, 信任率与几率水平(50%)的单样本 *t* 检验表明被试选择信任的比例显著高于几率水平, $t(20) = 5.61$, $p < 0.001$。选择信任和不信任的平均反应时分别为 495 ± 32 ms 和 490 ± 27 ms, 配对 *t* 检验表明两种决策的反应时无显著差异, $t(20) = 0.34$, $p = 0.736$。对不同反馈后信任选择率进行单因素重复测量方差分析, 结果发现不同反馈后信任选择率的主效应边缘显著, $F(2,40) = 3.21$, $p = 0.051$, $\eta^2 = 0.14$。事后比较发现获益反馈后信任选择率(60.6 ± 5.3%)显著低于中性反馈后信任选择率(71.8 ± 3.5%), $p < 0.05$, 损失反馈后信任选择率(68.9 ± 3.8%)与获益或中性反馈后信任选择率无显著差异, $ps > 0.05$。

3.2　时域 ERP 结果

P2　对 P2 (150~250 ms)最大峰值进行 2(决策

类型)×5(电极)的二因素重复测量方差分析，统计结果表明决策类型与电极位置的交互作用显著，$F(4,80) = 3.95$, $p = 0.036$, $\eta^2 = 0.17$。简单效应分析发现中央区(Cz)上不信任决策诱发的 P2 峰值(7.08 ± 1.01 μV)显著小于信任决策(8.11 ± 0.86 μV), $F(1,20) = 4.54$, $p = 0.046$；中央−顶区(CPz)上不信任决策诱发的 P2 峰值(6.40 ± 1.03 μV)显著小于信任决策(7.74 ± 1.03 μV), $F(1,20) = 6.23$, $p = 0.021$；后顶区(Pz)上不信任决策诱发的 P2 峰值(6.28 ± 1.19 μV)显著小于信任决策(7.84 ± 1.14 μV), $F(1,20) = 11.73$, $p = 0.003$；额区(Fz)和额−中央区(FCz)上两种决策类型诱发的 P2 无显著差异, $ps > 0.05$。

N2 对 N2(250~330 ms)最小峰值进行 2(决策类型)×5(电极)的二因素重复测量方差分析，统计结果表明决策类型的主效应显著, $F(1,20) = 5.61$, $p = 0.028$, $\eta^2 = 0.22$，表现为不信任决策诱发的 N2 峰值(1.44 ± 1.05 μV)显著小于信任决策(3.08 ± 0.85 μV)。电极位置前中后的主效应显著, $F(1,20) = 3.20$, $p = 0.017$, $\eta^2 = 0.14$，表现为 N2 峰值从前向后依次递增，额区(Fz)的 N2 波峰最小(1.35 ± 0.70 μV)，中央−顶区(CPz)的 N2 波峰最大(2.96 ± 1.09 μV)。

P3 对 P3(330~430 ms)最大峰值进行 2(决策类型)×5(电极)的二因素重复测量方差分析，统计结果表明决策类型的主效应边缘显著，$F(1,20) = 4.09$, $p = 0.057$, $\eta^2 = 0.17$，表现为不信任决策诱发的P3 峰值(6.60 ± 0.93 μV)边缘显著小于信任决策(7.66 ± 0.68 μV)。电极位置前中后的主效应显著, $F(1,20) = 3.51$, $p = 0.048$, $\eta^2 = 0.15$，表现为 P3 峰值从前向后依次递增，额区(Fz)的 P3 波峰最小(5.98 ± 0.55 μV)，中央−顶区(CPz)的 P3 波峰最大(7.77 ± 0.87 μV)。

3.3 频域 ERSP 结果

3.3.1 决策类型的 ERSP 结果

β 频段 对 β 频段能量(18~28 Hz, 250~350 ms)进行 2(决策类型)×5(电极)的二因素重复测量方差分析，结果发现决策类型的主效应显著，$F(1,20) = 5.67$, $p = 0.027$, $\eta^2 = 0.22$。事后配对比较发现不信任决策条件下 β 频段能量(−0.35 ± 0.19 dB)显著高于信任决策条件(−1.36 ± 0.33 dB)。其他主效应和交互作用不显著。

3.3.2 不同反馈后决策类型的 ERSP 结果

β 频段 对 β 频段能量(15~20 Hz, 150~300 ms)进行 2(反馈效价)×2(决策类型)×5(电极)的三因素重复测量方差分析，结果发现反馈效价与决策类型的交互作用边缘显著, $F(1,20) = 3.78$, $p = 0.066$, $\eta^2 = 0.16$。简单效应分析发现损失反馈后信任决策诱发的 β 频段能量(−0.72 ± 0.23 dB) 边缘显著大于获益反馈后信任决策诱发的 β 频段能量(−1.28 ± 0.27 dB), $F(1,20) = 4.13$, $p = 0.056$；损失反馈后不信任决策诱发的 β 频段能量(−0.97 ± 0.28 dB)与获益反馈后不信任决策诱发的 β 频段能量(−0.78 ± 0.23 dB)无显著差异, $F(1,20) = 0.56$, $p = 0.461$。其他主效应和交互作用不显著。

4 讨论

4.1 P2 反映了决策博弈中的风险评估

在决策选项呈现后 200 ms 左右，信任选择比不信任选择诱发了更大的 P2 成分。作为风险评估的早期成分，有研究认为 P2 与决策情境中风险性感知与评估有关，反映了决策启动的评估过程(Gajewski, Stoerig, & Falkenstein, 2008)。在人际互动任务中，P2 同样反映了个体对决策情境不确定性的感知，表现为高人际控制感条件下博弈决策诱发更小的 P2 成分(刘耀中, 窦凯, 2015)。当前研究采用单次匿名博弈范式，每一回合被试均需要与不同玩家完成博弈任务，因此信任选择具有较高的风险性与不确定性，进而导致更大的 P2 波幅。与本研究结果相一致，Polezzi 等发现相比于评估确定性方案而言，个体在评估不确定性方案时存在更大的 P2 波幅，并将这种结果解释为个体在不确定性方案上付诸了更多的认知资源(Polezzi, Lotto, Daum, Sartori, & Rumiati, 2008)。因此, P2 成分反映了博弈决策中的风险评估，单次匿名互动情境所具备的高风险性与不确定性使得信任选择诱发了更大的 P2 波幅。

4.2 N2 反映了决策博弈中的认知控制

在决策选项呈现后 300ms 左右，不信任选择比信任选择诱发了更负的 N2 波峰。作为冲突检测与认知控制的核心成分，N2 对冲突检测和反应抑制特别敏感，高冲突情境相比于低冲突情境能诱发更负的 N2 成分(Folstein & van Petten, 2008;van Veen & Carter, 2002; Schmajuk et al., 2006)。例如, Schmajuk 等(2006)采用 Stop-Signal 任务发现 N2 成分与冲突加工的正向关系，成功抑制相比于抑制失败能够诱发更强的右侧额叶 N2 成分。当前研究发现个体的信任行为显著高于几率水平，表明遵循强制规范的信任行为具有较高的优势性，因此选择不信任行为可能会引起较大的认知冲突与控制，进而诱发更负的 N2 成分。与当前研究结果相一致，刘耀中和窦凯(2015)发现低人际控制感条件下博弈决策诱发了

更大的 N2 波幅, 表明人际控制感越低, 被试在混合动机情境下所体验到的认知冲突越高。目前研究者关于信任或不信任何者为优势反应的理解仍存在较大争论, 有些研究者认为背叛厌恶动机驱动个体选择不信任以回避背叛所导致的负性效用, 另一些研究者则强调内化的道德规范会驱使人们选择信任以向他人品格表示尊重。当前研究中时域 N2 成分动态分离的结果则与强制规范理论相一致, 即决策博弈过程中人们更多地选择信任, 且信任决策为默认优势选项。与这种解释相一致, 新近有研究者收集并分析了 1400 名个体参与多种合作性博弈任务的行为决策, 结果发现不同合作性博弈中个体的合作行为高度相关, 进而证实人类社会中存在一种普遍存在且稳定的合作性表型(cooperative phenotype) (Peysakhovich, Nowak, & David, 2014)。综上所述, 信任博弈中个体做出不信任选择时会体验到较强的认知冲突, 需要消耗较多的认知资源来解决这种冲突, 最终诱发了更负的 N2 成分。

4.3　P3 反映了决策博弈中的晚期资源分配

在决策选项呈现后 400 ms 左右, 信任选择比不信任选择诱发了更大的 P3 波峰。作为一种晚期成分, P3 与刺激信号的识别、难度、任务、注意、记忆等认知心理过程有密切关系(Polich, 2007), 多数研究者认为 P3 主要反映了注意资源分配以及高水平的动机/情感评价(Leng & Zhou, 2010; Wang et al., 2013, 2014), 因此当前研究结果表明个体在信任选择上付诸了更多的注意资源。与当前研究结果相一致, Boudreau 等(2009)操纵了抛硬币任务中个体与匿名报告者利益冲突、利益一致和撒谎惩罚三种条件, 发现利益一致条件下报告者的声明信息诱发了更大的 P3 成分。因此, P3 成分反映了博弈决策中的晚期资源分配, 符合内化道德规范的信任选择具有更高水平的动机/情感意义, 能够导致更大的 P3 波幅。

4.4　β 频谱能量参与博弈决策中认知控制加工

决策类型的频域分析结果发现不信任决策比信任决策引发了更大的 β 频段(18~28 Hz, 250~350 ms)能量。早期研究发现静息状态下 β 频段活动往往集中于大脑运动系统上, 一度被认为是运动系统的闲置节律(idling rhythm)。新近研究者逐渐发现额–中央区域上的 β 震荡活动与认知控制加工有关, 认为它是另一种反映控制加工的有效生理指标(Aron, 2011; Huster, Enriquez-Geppert, Lavallee, Falkenstein, & Herrmann, 2013)。例如, Marco-Pallarés 等(2008)发现, 相比于停止信号任务中的反应试次和失败的抑制试次, 成功的抑制试次导致了更强的 β 频段能量。Krämer 等(2011)发现停止信号范式中成功抑制确实能够导致增强的 β 活动, 且这种效应在中央区域最大。因此, 不同决策类型诱发的 β 频段能量差异表明不信任选择比信任选择涉及到更强的认知控制加工, 是一种消耗更多认知资源的非优势选项。与当前研究结果相一致, Swann 等(2009)采用颅内 EEG 记录技术, 发现停止信号范式中刺激呈现后 100 至 250 ms 之间, 成功抑制试次能够导致右侧额下回上 β 频段活动增加。随后 Swann 等(2012)再次发现了成功抑制条件能够导致 β 活动的增强, 且颅内电极记录在右侧额下回皮层和前运动区均发现了上述效应。当前研究者关于人们为什么会信任或不信任陌生他人及何者为优势反应的理解仍存在较大争论。背叛厌恶理论强调背叛厌恶动机的普适性与重要性, 认为个体倾向于选择不信任以回避背叛所导致的负性效用; 而道德规范理论则认为人们凭借信任决策向他人品格表示尊重, 常常表现出较高的信任行为。当前研究的神经震荡结果则进一步支持了强制规范理论的观点, 即决策博弈过程中人们更多地选择信任, 且信任决策为默认优势选项。

另外, 先前结果能够调控随后行为决策及其诱发的 β 频段(15~20 Hz, 150~300 ms)能量, 表现为获益反馈后信任选择率显著低于中性反馈后信任选择率; 损失反馈后信任决策比获益反馈后信任决策诱发了更大的 β 频段能量, 而不同反馈后不信任决策诱发的 β 频段能量无显著差异。与本研究结果相一致, 有研究表明决策者在金钱获益后更倾向于规避风险, 而在金钱损失后更倾向于寻求风险(Gu et al, 2010; Polezzi et al., 2008)。频域分析结果则进一步说明, 在遭遇信任背叛之后, 决策者再次做出信任选择时会体验到更强烈的认知冲突, 需要消耗较多的认知资源来解决此冲突; 而在遭遇信任互惠后, 决策者再次做出信任选择时则体验到较少的上述认知冲突。

4.5　时域 ERP 和频域 ERO 分析方法相结合的优势

传统的脑电实验往往侧重于时域信号分析, 凭借叠加平均大量同质性试次事件相关的 EEG 信号来获得锁时锁相的 ERP, 并依据分段时窗内波形图上脑电成分的差异性分离来推测其间的大脑认知加工活动。这种时域 ERP 分析方法的优势体现为毫秒级的时间分辨率, 能够精确地刻画大脑认知加

工的动态时间进程，特别适宜于解答认知活动何时分离的问题(Wang et al., 2013, 2014)。然而，大脑的认知活动是以神经细胞群的电位能量变化为基石的，时域 ERP 仅仅反映了脑电信号中的非节律性活动或者锁时又锁相的震荡活动，却忽视了非严格锁相的震荡活动，因此无法阐明认知活动分离何以实现的问题。

相较而言，频域震荡分析使用傅里叶转换或小波分析等信号处理方法，实现了单试次 EEG 信号由时域向频域的转换，最终平均多个试次的频域信号从而得到事件相关的能量变化(ERO)，更适宜于从神经震荡水平揭示认知活动分离如何实现的问题(陈煦海，杨晓虹，杨玉芳, 2013)。联合采用 ERP 和 ERO 方法分析脑电数据不仅能精确地描绘不同认知活动动态分离的时间进程，而且可以从神经细胞群突触后电位的能量变化上揭示不同认知活动动态分离的实现途径，能够更清晰明朗地揭示心理认知过程的神经生理机制。以当前研究为例，时域 N2 成分和频域 β 频段能量的动态分离能够较好地解释信任与不信任陌生他人何者为优势反应/默认选项的问题，即在决策选项呈现后 250~350 ms 期间，不信任决策比信任决策诱发了显著更负的时域 N2 成分与更强的 β(18~28 Hz)频段能量，且两种指标均反映了额叶的认知控制功能，因此更强有力地证明了信任决策比不信任决策涉及更少的认知控制加工，属于信任博弈中的默认优势选项。上述研究结果不仅对已有信任博弈的研究成果进行了补充与扩展(Boudreau, et al., 2009; Wang et al., 2015)，而且进一步澄清了博弈决策中不同选择的优势性问题，为道德规范理论提供了一定的实证证据支持。

参 考 文 献

Aimone, J. A., Houser, D., & Weber, B. (2014). Neural signatures of betrayal aversion: An fMRI study of trust. *Proceedings of the Royal Society B: Biological Sciences, 281*(1782), 20132127.

Alegre, M., Gurtubay, I. G., Labarga, A., Iriarte, J., Valencia, M., & Artieda, J. (2004). Frontal and central oscillatory changes related to different aspects of the motor process: A study in go/no-go paradigms. *Experimental Brain Research, 159*(1), 14−22.

Aron, A. R. (2011). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biological Psychiatry, 69*(12), e55−e68.

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron, 58*(4), 639−650.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior, 10*(1), 122−142.

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization, 55*(4), 467−484.

Bohnet, I., Greig, F., Herrmann, B., & Zeckhauser, R. (2008). Betrayal aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States. *American Economic Review, 98*(1), 294−310.

Boudreau, C., McCubbins, M. D., & Coulson, S. (2009). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. *Social Cognitive and Affective Neuroscience, 4*(1), 23−34.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. New York: Princeton University Press

Chen, J., Zhong, J., Zhang, Y. X., Li, P., Zhang, A. Q., Tan, Q. B., & Li, H. (2012). Electrophysiological correlates of processing facial attractiveness and its influence on cooperative behavior. *Neuroscience Letters, 517*(2), 65−70.

Chen, X. H., Yang, X. H., & Yang, Y. F. (2013). Neurophysiological mechanism of implicit processing of vocal emotion transition. *Acta Psychologica Sinica, 45*(4), 416−426.

[陈煦海，杨晓虹，杨玉芳. (2013). 语音情绪变化内隐加工的神经生理机制. *心理学报, 45*(4), 416−426.]

Declerck, C. H., Boone, C., & Emonds, G. (2013). When do people cooperate? The neuroeconomics of prosocial decision making. *Brain and Cognition, 81*(1), 95−117.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9−21.

Dunning, D., Anderson, J. E., Schlösser, T., Ehlebracht, D., & Fetchenhauer, D. (2014). Trust at zero acquaintance: More a matter of respect than expectation of reward. *Journal of Personality and Social Psychology, 107*(1), 122–141.

Dunning, D., Fetchenhauer, D., & Schlösser, T. M. (2012). Trust as a social and emotional act: Noneconomic considerations in trust behavior. *Journal of Economic Psychology, 33*(3), 686–694.

Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152−170.

Gajewski, P. D., Stoerig, P., & Falkenstein, M. (2008). ERP—correlates of response selection in a response conflict paradigm. *Brain Research, 1189*, 127−134.

Gu, R. L., Ge, Y., Jiang, Y., & Luo, Y. J. (2010). Anxiety and outcome evaluation: the good, the bad and the ambiguous. *Biological Psychology, 85*(2), 200−206.

Huster, R. J., Enriquez-Geppert, S., Lavallee, C. F., Falkenstein, M., & Herrmann, C. S. (2013). Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. *International Journal of Psychophysiology, 87*(3), 217−233.

Johnson, N. D., & Mislin, A. A. (2011). Trust games: A meta-analysis. *Journal of Economic Psychology, 32*(5), 865−889.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science, 308*(5718), 78−83.

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of*

the National Academy of Sciences of the United States of America, 106(49), 20895−20899.
Krämer, U. M., Knight, R. T., & Münte, T. F. (2011). Electrophysiological evidence for different inhibitory mechanisms when stopping or changing a planned response. *Journal of Cognitive Neuroscience, 23*(9), 2481−2493.
Krueger, F., Grafman, J., & McCabe, K. (2008). Neural correlates of economic game playing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1511), 3859−3874.
Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M.,... Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences of the United States of America, 104*(50), 20084−20089.
Leng, Y., & Zhou, X. L, (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia, 48*(2), 448−455.
Liu, Y. Z., & Dou, K. (2015). The effect of perceived interpersonal control on cooperation behavior: An ERP investigation using the Prisoner’s Dilemma Game. *Journal of Psychological Science, 38*(3), 643−650.
[刘耀中, 窦凯. (2015). 人际控制感对合作行为的影响: 一项囚徒困境的 ERPs 研究. *心理科学, 38*(3), 643−650.]
Long, Y., Jiang, X., & Zhou, X. (2012). To believe or not to believe: Trust choice modulates brain responses in outcome evaluation. *Neuroscience, 200*, 50−58.
Ma, Q. G., Meng, L., & Shen, Q. (2015). You have my word: Reciprocity expectation modulates feedback-related negativity in the Trust Game. *PLoS One, 10*(2), e0119129.
Marco-Pallarés, J., Camara, E., Münte, T. F., & Rodríguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of Cognitive Neuroscience, 20*(9), 1595−1610.
Masaki, H., Takeuchi, S., Gehring, W. J., Takasawa, N., & Yamazaki, K. (2006). Affective-motivational influences on feedback-related ERPs in a gambling task. *Brain Research, 1105*(1), 110−121.
McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences of the United States of America, 98*(20), 11832−11835.
Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 17−26.
Peysakhovich, A., Nowak, M. A., & Rand, D. G. (2014). Humans display a ‘cooperative phenotype’ that is domain general and temporally stable. *Nature Communications, 5*, 4939.
Polezzi, D., Lotto, L., Daum, I., Sartori, G., & Rumiati, R. (2008). Predicting outcomes of decisions in the brain. *Behavioural Brain Research, 187*(1), 116−122.
Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical neurophysiology, 118*(10), 2128−2148.
Riedl, R., & Javor, A. (2012). The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics, 5*(2), 63−91.
Schmajuk, M., Liotti, M., Busse, L., & Woldorff, M. G. (2006). Electrophysiological activity underlying inhibitory control processes in normal adults. *Neuropsychologia, 44*(3), 384−395.
Swann, N., Tandon, N., Canolty, R., Ellmore, T. M., McEvoy, L. K., Dreyer, S., ... Aron, A. R. (2009). Intracranial EEG reveals a time-and frequency-specific role for the right inferior frontal gyrus and primary motor cortex in stopping initiated responses. *Journal of Neuroscience, 29*(40), 12675−12685.
Swann, N. C., Cai, W. D., Conner, C. R., Pieters, T. A., Claffey, M. P., George, J. S., ... Tandon, N. (2012). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: electrophysiological responses and functional and structural connectivity. *NeuroImage, 59*(3), 2860−2870.
Tiesinga, P., Fellous, J. M., & Sejnowski, T. J. (2008). Regulation of spike timing in visual cortical circuits. *Nature Reviews Neuroscience, 9*(2), 97−107.
Tzieropoulos, H. (2013). The Trust Game in neuroscience: A short review. *Social Neuroscience, 8*(5), 407−416.
van Veen, V., & Carter, C. S. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience, 14*(4), 593−602.
Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews, 90*(3), 1195−1268.
Wang, Y. W., Roberts, K., Yuan, B., Zhang, W. X., Shen, D. L., & Simons, R. (2013). Psychophysiological correlates of interpersonal cooperation and aggression. *Biological Psychology, 93*(3), 386−391.
Wang, Y. W., Yuan, B., Roberts, K., Wang, Y., Lin, C. D., & Simons, R. F. (2014). How friendly is a little friendly competition? Evidence of self-interest and empathy during outcome evaluation. *International Journal of Psychophysiology, 91*(3), 155−162.
Wang, Y. W., Zhang, Z., Yuan, S., Guo, F. B., He, S. Y., & Jing, Y. M. (2015). The decision-making and outcome evaluation during a repeated Trust Game. *Acta Psychologica Sinica, 47*(8), 1028−1038.
[王益文, 张振, 原胜, 郭丰波, 何少颖, 敬一鸣. (2015). 重复信任博弈的决策过程与结果评价. *心理学报, 47*(8), 1028−1038.]
Wiswede, D., Münte, T. F., & Rüsseler, J. (2009). Negative affect induced by derogatory verbal feedback modulates the neural signature of error detection. *Social Cognitive and Affective Neuroscience, 4*(3), 227−237.
Zak, P. J., & Knack, S. (2001). Trust and growth. *The Economic Journal, 111*(470), 295−321.

Brain dynamics of decision-making in the generalized trust game: Evidence from ERPs and EEG time-frequency analysis

FU Chao[1,2]; ZHANG Zhen[1]; HE Jinzhou[3]; HUANG Silin[4]; QIU Jianyin[5]; WANG Yiwen[1,2]

([1] *Institute of Psychological and Cognitive Sciences, Fuzhou University, Fuzhou* 350116, *China*)

([2] *School of Economics and Management, Fuzhou University, Fuzhou* 350116, *China*)

([3] *Department of Psychology, Chung Yuan Christian University, Taoyuan* 32023, *China*)

([4] *Institute of Developmental Psychology, Beijing Normal University, Beijing* 100875, *China*)

([5] *Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai* 200030, *China*)

Abstract

Human societies are unique in terms of large-scale social cooperation and trust between unrelated people (generalized trust). As an important social signaling mechanism, generalized trust reduces transaction costs, facilitates cooperative behavior, and promotes human society's prosperity. The question of why and when people trust each other has important implications for human social interaction and economic decision-making. Previous fMRI research utilizing the Trust Game has revealed brain regions involved in the decision to trust, including medial prefrontal cortex, caudate nucleus, amygdala, insula and dorsolateral prefrontal cortex. However, the extant research has failed to clearly reveal neurophysiological mechanisms of the investor's decision-making in the Trust Game.

To investigate the time course and neural oscillation courses of trust decision-making, we recorded the electroencephalographic (EEG) data of 21 healthy human participants while they played the role of trustor in the one-shot Trust Game; in this game, participants made decisions to trust or distrust the counterpart. In addition to behavioral data analysis and event-related potential (ERP) analysis, we also conducted spectral analysis to examine the neural oscillations underlying the trust game responses.

Behavioral results indicated that participants made trusting decisions more than chance. Electrophysiological results suggested that distrusting choices induced a more negative N2 and a less positive P3 than did trusting choices. Spectral analysis results showed that the beta-band (18~28 Hz, 250~350 ms) power for distrusting choices was significantly larger than trusting choices; this may suggest greater inhibitory control exerted to distrust. Moreover, trusting choice after a negative outcome elicited a significant higher beta-band (15~20 Hz, 150~300 ms) power than did trusting choice after a positive outcome, while the beta-band power for distrusting choice after a negative outcome was similar compared to distrusting choice after a positive outcome.

Our study provided some new insights into the psychophysiological processes underlying human's decisions to trust strangers. Specifically, when making a decision to trust or distrust the counterpart, the distrusting choice induced a larger N2 than did the trusting choice, which may reflect greater cognitive control to inhibit trusting behavior. On the other hand, a more positive going P3 for the trusting choice than the distrusting choice may suggest that increased attention and motivation are associated with the trusting behavior. Moreover, the increased beta power across frontal areas for distrusting choices might reflect frontal inhibitory processes. These findings highlight the ingrained norm of cooperation and trust in modern societies, and provide preliminary spatio-temporal and spectral results understanding human's decision-making in the one-shot Trust Game.

Key words generalized trust; decision-making; N2; P3; Time-frequency analysis